

BioVEC: A program for Biomolecule Visualization with Ellipsoidal Coarse-graining

Erik Abrahamsson*, Steven S. Plotkin†

Department of Physics and Astronomy,
The University of British Columbia,
6224 Agricultural Road, Vancouver, B.C. V6T 1Z1, Canada

Accepted for publication in Journal of Molecular Graphics and Modelling

doi:10.1016/j.jmgm.2009.05.001

May 18, 2009

Biomolecule Visualization with Ellipsoidal Coarse-graining (BioVEC) is a tool for visualizing molecular dynamics simulation data while allowing coarse-grained residues to be rendered as ellipsoids. BioVEC reads in configuration files, which may be output from molecular dynamics simulations that include orientation output in either quaternion or ANISOU format, and can render frames of the trajectory in several common image formats for subsequent concatenation into a movie file. The BioVEC program is written in C++, uses the OpenGL API for rendering, and is open source. It is lightweight, allows for user-defined settings for and texture, and runs on either Windows or Linux platforms.

1 Introduction

The use of coarse-graining methods in biomolecular simulation is progressively increasing [1], mainly to circumvent the current time-scale problem that exists for all-atom molecular dynamics simulations, and to allow for the computational modeling of ever larger systems which address biologically relevant questions. However as the coarse-grained constituents encompass more atoms, the resulting residual objects become poorly approximated by spheres with corresponding isotropic potentials, while the steric character of the residues becomes more important. Such a scenario would exist for the coarse-graining of bases in DNA, or the coarse-graining of helices in a protein.

For these reasons, interest in ellipsoidal coarse-graining, perhaps the simplest non-spherical generalization, has recently increased [2–9]. Interest in industrial applications such as the manufacture of nanoparticle films has also motivated studies of ellipsoidal coarse-graining [10, 11].

The molecular dynamics (MD) of rigid objects raises some new practical challenges, including a

singularity-free representation of rigid body rotation [12–15] and corresponding energy conserving scheme for symplectic integration [16–18]. This is best done using quaternions [19, 20], which MD simulation programs such as LAMMPS [7] currently allow one to employ.

Quite apart from the issues of coarse-graining and molecular dynamics, visualization tools for rendering thermal ellipsoids representing Debye-Waller factors have enjoyed a long and successful tradition [21]. However, in spite of widely used standards in crystallography, as well as the above-mentioned progress in coarse-graining and molecular dynamics of rigid bodies, to the authors' knowledge no currently available program allows for the straightforward visualization of the dynamical evolution of coarse-grained systems with ellipsoidal constituents. To facilitate our own analysis of such coarse-grained models, we have decided to develop our own program. The program is portable to either Windows or Linux systems, and could in principle be incorporated into many of the more versatile and widely-used visualization programs such as VMD [22], Molmol [23], Molscript [24], PyMOL [25], Raster3D [26], Jmol [27], and Rasmol [28].

*E-mail: erik@phas.ubc.ca

†E-mail: steve@phas.ubc.ca

2 Methods

The program BioVEC presented in this paper, is a program to visualize the results of MD simulations of biomolecules that have been coarse-grained with ellipsoids. The functionality of BioVEC is simple and basic, as its purpose is not to compete with current molecular visualization packages but specifically to allow the scientist to visualize dynamical processes involving ellipsoidal coarse-graining. BioVEC automatically creates graphical images, which can be compiled to movie files in an external editor, such as Slide Show Movie Maker [29] (Windows) or MJPEG Tools [30] (*NIX). It is also possible for the user to manually step through the simulation, and to save the current view of the molecule to an image file in several formats, including `png`, `bmp`, and `jpg`. The representation used in BioVEC differs from techniques such as cartoon and ribbon representation, as these techniques aim to smooth out details, in all-atom simulations for example. The aim of BioVEC is rather to show all the features of already coarse-grained simulation data.

BioVEC reads in molecular dynamics simulation data from a file containing the positions of the coarse-grained species and the orientations of the ellipsoids for all time steps in the molecular dynamics simulation. The input format for BioVEC is based on the output dump file format in LAMMPS [7], which is of the form:

```
species-no  species-type  position × 3
orientation × 4
```

for each time step. The program also reads in a file with information on the topology of the molecule, in order for the program to display bonds between the species. In the current version of the program, covalent bonds are not allowed to break or form between different time steps. The information in the topology file comes from the LAMMPS input file, and is of the form

```
bond-no bond-type 1st-species 2nd-species
1st-origin 2nd-origin
```

where the bond goes from the `1st-origin` on the `1st-species` to the `2nd-origin` on the `2nd-species`, corresponding to the `species-no` in the simulation data file. Each ellipsoid can have an arbitrary number of bond origins, as specified in the input file. The `1st-origin` and `2nd-origin` entries in the topology file are optional if the species only have one bond origin each.

The user supplies the names of the simulation data and topology files in an input file with information about the coarse-graining of the system. This file should contain the number of coarse-grained species, the radii of the different ellipsoids and spheres, as well

as the vector from the centre of mass of the ellipsoid to the location of the bond connecting the ellipsoidal residue to its covalently-bonded partner. Spheres are treated as ellipsoids with all three radii equal. Information about the colouring and texture of the species is also entered into the input file.

The BioVEC program can also read and visualize ANISOU anisotropic temperature factor records in PDB standard form [31].

2.1 Quaternions

Quaternions [19] are frequently used in computer graphics to represent rotations and orientation in 3-dimensional space. One of the main advantages of quaternions over other representations, such as Euler angles, is the absence of singularities in the quaternion representation.

A quaternion is a four-dimensional extension to the complex numbers,

$$\mathbf{q} = q_0\mathbf{u} + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}. \quad (1)$$

where q_0 , q_1 , q_2 , and q_3 are real numbers, and \mathbf{u} , \mathbf{i} , \mathbf{j} , and \mathbf{k} are such that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{i}\mathbf{j}\mathbf{k} = -\mathbf{u}^2 = -1$, and $\mathbf{i}\mathbf{j} = -\mathbf{j}\mathbf{i} = \mathbf{k}$. Quaternions are often written in a more compact vector form,

$$\mathbf{q} = [q_0, q_1, q_2, q_3], \quad (2)$$

where the quantities q_0 , q_1 , q_2 , and q_3 again are real numbers.

The relation between the quaternion representation and the rotation and angle α about a unit axis, $\hat{\mathbf{n}} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ is given by (see for example [32])

$$\begin{aligned} s &= (q_1^2 + q_2^2 + q_3^2)^{1/2}, \\ x &= q_1/s, \\ y &= q_2/s, \\ z &= q_3/s, \\ \alpha &= 2\cos^{-1}(q_0), \end{aligned} \quad (3)$$

for $s \neq 0$. Note that for the q_i to represent a rotation $\sum_{i=0}^3 q_i^2 = 1$. The orientation of the different ellipsoids are read by BioVEC from the simulation dump-file in vector form (Eq. 2), together with the position and ellipsoid type.

2.2 Bond vectors

The bonds between ellipsoidal species are drawn between the same points on the ellipsoids as were used in the MD simulation. The vectors from the centre of mass of the different types of ellipsoids to the location

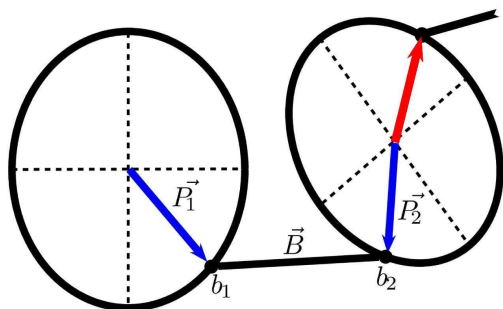


Figure 1: Representation of the bond vectors. The bond \vec{B} is formed between the points b_1 and b_2 , given by the bond vectors \vec{P}_1 and \vec{P}_2 . A second bond vector for the right ellipsoid is indicated by a red arrow. The bond vectors are given relative to the principal axis (dotted lines) of the ellipsoids.

where the covalent bond joins the ellipsoid (Fig. 1), are supplied by the user in the configuration file. The vectors are given in the principle (body-centered) basis of the ellipsoid for each bond origin on each ellipsoid species.

3 Implementation

The BioVEC program is written in C++, using the OpenGL API for rendering. The GLUT library is used for creating menus. The user input is command line driven, and the interaction with the program is via keyboard commands and mouse menus.

Graphical images from the program can be saved to file in several different formats through the DevIL library [33]. Images can be saved automatically for each time step of the simulation as the program reads through the data file, or the current view of the molecule can be saved by the user.

The BioVEC source can be compiled and run on both the Windows and Linux platforms.

4 Results and Discussion

Figure 2 shows an example of BioVEC running under Windows XP, depicting the console window, the OpenGL window, and the GLUT mouse menu. The molecule can be rotated and zoomed with either the keyboard or the mouse, and the time stepping can be done automatically or one step at a time by the user. The user can let the program automatically print the current view to an image file, or can choose to manually print the current view with a press of a key. This

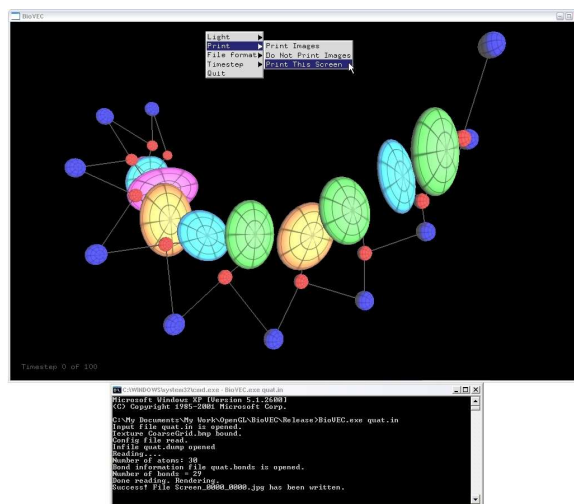


Figure 2: View of the BioVEC interface, running on Windows XP.

way stereo views of the molecule can be created.

The advantage of using a program, such as BioVEC, capable of visualizing ellipsoids, is seen in Fig. 3, where the all-atom model of single-stranded DNA is shown alongside with a coarse-grained ball-and stick model, and an ellipsoidal model, both visualized in BioVEC. In the all-atom model the orientation of the bases can be seen, while this information is lost when displaying the coarse-grained DNA strand with the ball-and stick model. The ellipsoidal visualization model used in BioVEC, however, retains the orientational information of the all-atom representation.

Figure 4 shows a collapsed globule of single-stranded DNA with block-copolymeric sequence A($\times 15$)- T($\times 15$). Information that would not have been seen with a coarse-grained ball-and-stick rendering, such as partial hair-pin formation along with orientation of the stacked base pairs, is clearly visualized in BioVEC.

The input format of the program also allows for multiple bond origins, as shown in Fig. 5, where a 4-helix bundle is coarse-grained so that the helices are modeled by effective prolate ellipsoids. Investigating the folding mechanism of such a coarse-grained system allows one to investigate the assumptions of the diffusion-collision model [34], for example. Renderings with multiple bond origins may be useful for characterizing the constraints due to intermolecular chemical bonds between biomolecules.

Separate strands of coarse-grained polymers can be visualized in BioVEC. An example of this is shown in Fig. 6, where a polymer glass melt, consisting of 100 polymer strands with ten atoms in each strand [35], is visualized. By giving all the atoms on a single

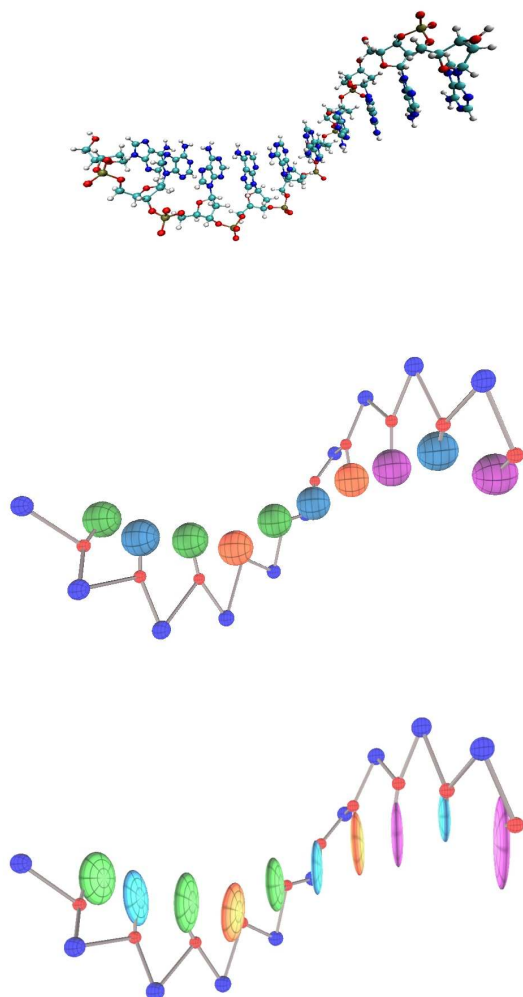


Figure 3: Comparison of three different visualization models for single-stranded DNA. All-atom CPK representation rendered in VMD; coarse-grained ball-and-stick model; and coarse-grained ellipsoidal model, both rendered in BioVEC.

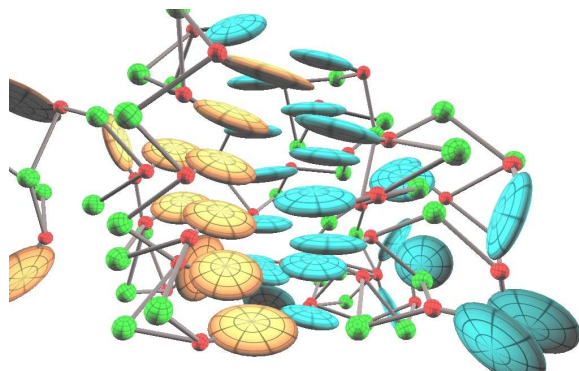


Figure 4: BioVEC visualization of single stranded DNA consisting of adenine and thymine, forming a hairpin.

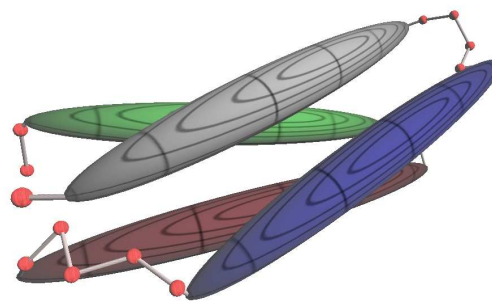


Figure 5: BioVEC visualization of the homodimeric ROP protein 1YO7 four helix bundle, with the helices represented by ellipsoids. Multiple bond origins are present for this system, and represent the points where the backbone chain enters and exits the helices.

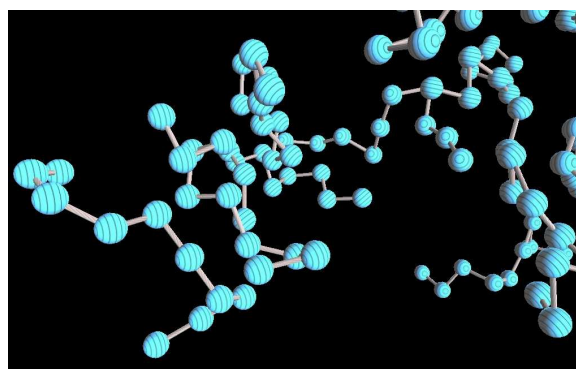


Figure 6: BioVEC visualization of polymer glass [35]. The simulation contains 100 polymers with ten atoms in each polymer. Each polymer is visualized with different orientation of the texture to simplify the interpretation of the image.

polymer the same orientation, the separate polymers in the glass can be identified. The information about the bonds, and thereby the separate polymers, is given in the topology file.

BioVEC can also read and visualize PDB files with ANISOU data. Figure 7 shows the thermal ellipsoid representation of GM1-pentasaccharide from a complex with the cholera toxin B-pentamer refined with anisotropic displacement parameters at 1.25 Å resolution (PDB data taken from a Raster3D [26] example file).

BioVEC can be used to visualize ellipsoids for a wide range of applications. Figure 8 shows the orientation of the dielectric tensors at a cross-section through the centre of ubiquitin [36]. The ellipsoids represents the local dielectric constant, with the orientation given by treating the dielectric tensor as a rotation matrix. The sizes of the principal axis are given by the reciprocals

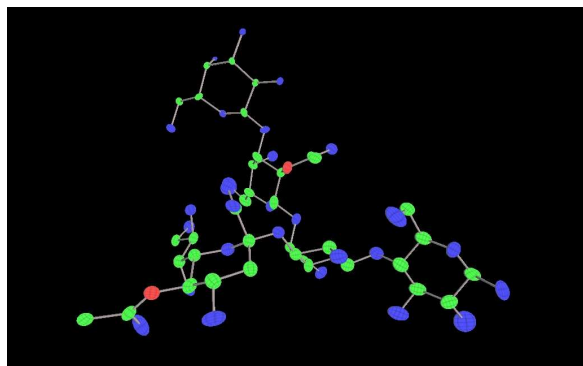


Figure 7: BioVEC visualization of the thermal anisotropy of GM1-pentasaccharide from a complex with the cholera toxin B-pentamer [26].

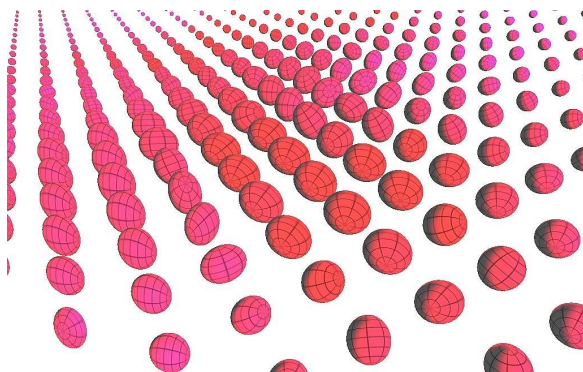


Figure 8: BioVEC visualization of the dielectric tensors at a cross-section of ubiquitin. [36]. The principal axis of the ellipsoids are given by the inverse of the eigenvalues of the dielectric tensor.

of the eigenvalues, to emphasize the low dielectric constant in the centre of the protein.

5 Conclusion

The BioVEC program presented here is a simple but useful tool for visualizing bio-molecules that are coarse-grained with ellipsoids. There are many program packages available for molecular visualization, but none of these can, to our knowledge, display biomolecules coarse-grained with ellipsoids. Some programs, such as Raster3D [26] and PyMOL [25], allow for the visualization of anisotropic temperature factors in the form of ellipsoids. However, they are currently not suitable for straight-forward visualization of the dynamics of coarse-grained ellipsoidal systems with the orientation represented by quaternions. The BioVEC program therefore focuses on this feature, making it a small and user-friendly program for visualization of coarse-grained molecular dynamics data.

Future revisions of BioVEC should include a graph-

ical user interface (GUI), in order to improve both the functionality and the ease-of-use. Improvement and optimization of the rendering algorithm should accelerate rendering of large molecules, or long simulation runs.

BioVEC is open source, and is distributed under the GNU GPL licence. The source code and Windows executable are available for download at <http://www.phas.ubc.ca/~steve/BioVEC/>.

6 Acknowledgments

E.A gratefully acknowledges support from the Sweden-America Foundation and the Olle Engkvist Foundation. S.S.P. gratefully acknowledges support from the Natural Sciences and Engineering Research Council and the A. P. Sloan Foundation. We are grateful to Alex Morriss-Andrews, Will Guest, and Mya Warren for supplying simulation data and for helpful discussions.

References

- [1] V. Tozzini, Coarse-grained models for proteins, *Curr. Op. Str. Bio* 15 (2) (2005) 144–150.
- [2] J. G. Gay, B. J. Berne, Modification of the overlap potential to mimic a linear site-site potential, *J. Chem. Phys.* 74 (1981) 3316–3319.
- [3] R. Berardi, C. Fava, C. Zannoni, A Gay-Berne potential for dissimilar biaxial particles, *Chem. Phys. Lett.* 297 (1998) 8–14.
- [4] R. Everaers, M. R. Ejtehadi, Interaction potentials for soft and hard ellipsoids, *Phys. Rev. E* 67 (2003) 041710.
- [5] L. Paramonov, S. N. Yaliraki, The directional contact distance of two ellipsoids: Coarse-grained potentials for anisotropic interactions, *J. Chem. Phys.* 123 (2005) 194111.
- [6] M. Babadi, R. Everaers, M. R. Ejtehadi, Coarse-grained interaction potentials for anisotropic molecules, *J. Chem. Phys.* 124 (2006) 174708.
- [7] S. J. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comp. Phys.* (1995) 1–19.
- [8] R. Berardi, L. Muccioli, S. Orlandi, M. Ricci, C. Zannoni, Computer simulations of biaxial nematics, *J Physics CM* 20 (2008) 463101.
- [9] J. Stelzer, L. Longa, H.-R. Trebin, Molecular dynamics simulations of a Gay-Berne nematic liquid crystal: Elastic properties from direct correlation functions, *J. Chem. Phys.* 103 (1995) 3098–3107.
- [10] G. S. Grest, W. M. Brown, J. B. Lechman, M. K. Petersen, S. J. Plimpton, P. R. Schunk, Nanoparticle flow, ordering, and self-assembly, *SAND Reports* 6516 (2008) 1–29.
- [11] W. M. Brown, M. K. Petersen, S. J. Plimpton, G. S. Grest, Liquid crystal nanodroplets in solution, *J. Chem. Phys.* 130 (2009) 044901.

- [12] D. J. Evans, Representation of orientation space, *Mol. Phys.* 34 (1977) 317–325.
- [13] D. J. Evans, S. Murad, Singularity free algorithm for molecular-dynamics simulation of rigid polyatomics, *Mol. Phys.* 34 (1977) 327–331.
- [14] M. Nitschke, E. H. Knickmeyer, Rotation parameters - a survey of techniques, *J. Surv. Engrg.* 126 (2000) 83–105.
- [15] E. A. Coutias, C. Seok, K. A. Dill, Using quaternions to calculate RMSD, *J. Comput. Chem.* (2004) 1849–1857.
- [16] T. F. Miller III, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. News, G. J. Martyna, Symplectic quaternion scheme for biophysical molecular dynamics, *J. Chem. Phys.* 116 (2002) 8649–8659.
- [17] R. Shivarama, E. P. Farenthold, Hamilton’s equations with Euler parameters for rigid body dynamics modeling, *J. Dyn. Sys., Meas., Control* 126 (2004) 124–130.
- [18] H. Kamberaj, R. J. Low, M. P. Neal, Time reversible and symplectic integrators for molecular dynamics simulations of rigid molecules, *J. Chem. Phys.* 122 (2005) 224114.
- [19] W. R. Hamilton, On a new species of imaginary quantities connected with a theory of quaternions, *Rroc. R. Irish. Acad.* 2 (1844) 424–434.
- [20] C. F. F. Karney, Quaternions in molecular modelling, *J. Mol. Graph. Model.* 25 (2007) 595–604.
- [21] M. N. Burnett, C. K. Johnson, ORTEP-III: Oak Ridge Thermal Ellipsoid Plot program for crystal structure illustrations, Oak Ridge National Laboratory Report ORNL-6895.
- [22] W. Humphrey, A. Dalke, K. Schulten, VMD – Visual Molecular Dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [23] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures., *J Mol Graphics* 14 (1996) 51–55.
- [24] P. J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J Appl. Cryst.* 24 (1991) 946–950.
- [25] PyMOL Molecular Viewer, <http://www.pymol.org/> (2009).
- [26] E. A. Merritt, D. J. Bacon, Raster3d – Photorealistic molecular graphics, *Methods in Enzymology* 277 (1997) 505–524.
- [27] Jmol: an open-source java viewer for chemical structures in 3D, <http://www.jmol.org/> (2009).
- [28] RasMol: Molecular graphics visualisation tool, <http://www.rasmol.org/> (2009).
- [29] SSMM – Slide Show Movie Maker, <http://www.joern-thiemann.de/tools/SSMM> (2009).
- [30] MJPEG Tools, <http://mjpeg.sourceforge.net> (2009).
- [31] Pdb format, version 2.3, <http://www.wwpdb.org/documentation/format23/sect9.html#ANISOU> (2009).
- [32] E. W. Weisstein, “Quaternion”. From MathWorld – a Wolfram Web Resource, <http://mathworld.wolfram.com/Quaternion.html> (2008).
- [33] D. Woods, N. Weber, M. Dario, DevIL 1.7.2 – a full featured cross-platform image library, <http://openil.sourceforge.net> (2008).
- [34] M. Karplus, D. L. Weaver, Protein folding dynamics: the diffusion-collision model and experimental data, *Protein Science* 3 (1994) 650–668.
- [35] M. Warren, J. Rottler, Heterogeneous diffusion in amorphous solids, In press.
- [36] W. Guest, N. Cashman, S. S. Plotkin, On the anisotropic and inhomogeneous dielectric properties of proteins, Preprint.